



EP3260: Machine Learning Over Networks
Computer Assignment 1
Due Date: February 14, 2023

Computer Assignment 1 - Closed-form solution vs iterative approaches

Let us consider

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathbb{R}^d}{\text{minimize}} \frac{1}{N} \sum_{i \in [N]} \|\mathbf{w}^T \mathbf{x}_i - \mathbf{y}_i\|^2 + \lambda \|\mathbf{w}\|_2^2,$$

for a dataset $\{(\mathbf{x}_i, \mathbf{y}_i)\}$.

Then, address the following:

- (a) Find a closed-form solution for this problem;
- (b) Consider “Individual household electric power consumption” dataset ($N = 2075259$, $d = 9$) and find the optimal linear regressor from the closed-form expression;
- (c) Repeat 2) for “Greenhouse gas observing network” dataset ($N = 2921$, $d = 5232$) and observe the scalability issue of the closed-form expression;
- (d) How would you address even bigger datasets?